US009437204B2

(12) **United States Patent**
Grancharov et al.

(10) **Patent No.: US 9,437,204 B2**
(45) **Date of Patent: Sep. 6, 2016**

(54) **TRANSFORM ENCODING/DECODING OF HARMONIC AUDIO SIGNALS**

(71) Applicant: **Telefonaktiebolaget L M Ericsson (publ)**, Stockholm (SE)

(72) Inventors: **Volodya Grancharov**, Solna (SE); **Tomas Jansson Toftgård**, Uppsala (SE); **Sebastian Näslund**, Solna (SE); **Harald Pobloth**, Täby (SE)

(73) Assignee: **Telefonaktiebolaget LM Ericsson (publ)**, Stockholm (SE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 25 days.

(21) Appl. No.: **14/387,367**

(22) PCT Filed: **Oct. 30, 2012**

(86) PCT No.: **PCT/SE2012/051177**

§ 371 (c)(1),
(2) Date: **Sep. 23, 2014**

(87) PCT Pub. No.: **WO2013/147666**

PCT Pub. Date: **Oct. 3, 2013**

(65) **Prior Publication Data**

US 2015/0046171 A1 Feb. 12, 2015

Related U.S. Application Data

(60) Provisional application No. 61/617,216, filed on Mar. 29, 2012.

(51) **Int. Cl.**
G10L 19/00 (2013.01)
G10L 19/02 (2013.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G10L 19/0212*** (2013.01); ***G10L 19/002*** (2013.01); ***G10L 19/028*** (2013.01); ***G10L 19/038*** (2013.01)

(58) **Field of Classification Search**
CPC ... G10L 19/002; G10L 19/008; G10L 19/04; G10L 19/038; G10L 19/167; G10L 19/24; G10L 21/038; G10L 19/093; G10L 19/12; G10L 19/18; G10L 2019/0004; G10L 19/012; G10L 19/0208; G10L 19/0212; G10L 19/022; G10L 19/028; G10L 19/032; G10L 19/035; G10L 16/265; G10L 21/04; G10L 21/043; G10L 25/90
USPC 704/500, 200.1, 219, 230, 229, 200, 704/205, 208, 211, 214, 216, 222, 226–228, 704/501–504, 233, 203, 232, 206
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,263,312 B1 * 7/2001 Kolesnik G10L 19/0208 704/229
7,831,434 B2 * 11/2010 Mehrotra G10L 21/038 381/21

(Continued)

FOREIGN PATENT DOCUMENTS

RU 2436174 C2 12/2011
WO 2011063694 A1 6/2011
WO 2011114933 A1 9/2011

OTHER PUBLICATIONS

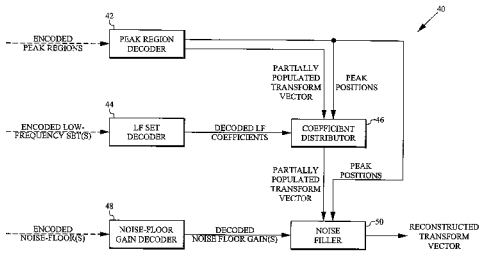
Bartkowiak, Maciej et al., "Harmonic Sinusoidal + Noise Modeling of Audio based on Multiple F0 Estimation", Audio Engineering Society, Convention Paper 7510, 125th Convention, San Francisco, CA, Oct. 2-5, 2008, 1-8.

Primary Examiner — Vijay B Chawan
(74) *Attorney, Agent, or Firm* — Murphy, Bilak & Homiller, PLLC

(57) **ABSTRACT**

An encoder (**20**) for encoding frequency transform coefficients (Y(k)) of a harmonic audio signal include the following elements: A peak locator (**22**) configured to locate spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold. A peak region encoder (**24**) configured to encode peak regions including and surrounding the located peaks. A low-frequency set encoder (**26**) configured to encode at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions. A noise-floor gain encoder (**28**) configured to encode a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

16 Claims, 12 Drawing Sheets



(51) **Int. Cl.**

G10L 19/002	(2013.01)
G10L 19/028	(2013.01)
G10L 19/038	(2013.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,885,819	B2 *	2/2011	Koishida	G10L 19/167 704/500
7,953,604	B2 *	5/2011	Mehrotra	G10L 21/038 704/200.1
8,046,214	B2 *	10/2011	Mehrotra	G10L 19/008 341/155
8,392,179	B2 *	3/2013	Yu	G10L 19/093 704/200
2007/0238415	A1	10/2007	Sinha et al.	
2008/0319739	A1 *	12/2008	Mehrotra	G10L 19/008 704/200.1
2011/0010168	A1 *	1/2011	Yu	G10L 19/093 704/219
2011/0035226	A1 *	2/2011	Mehrotra	G10L 19/008 704/500
2011/0178795	A1 *	7/2011	Bayer	G10L 19/002 704/205
2011/0196684	A1 *	8/2011	Koishida	G10L 19/167 704/500
2012/0029923	A1	2/2012	Rajendran et al.	
2012/0046955	A1	2/2012	Rajendran et al.	
2012/0259645	A1 *	10/2012	Budnikov	G10L 19/012 704/500
2012/0323584	A1 *	12/2012	Koishida	G10L 19/167 704/500

* cited by examiner

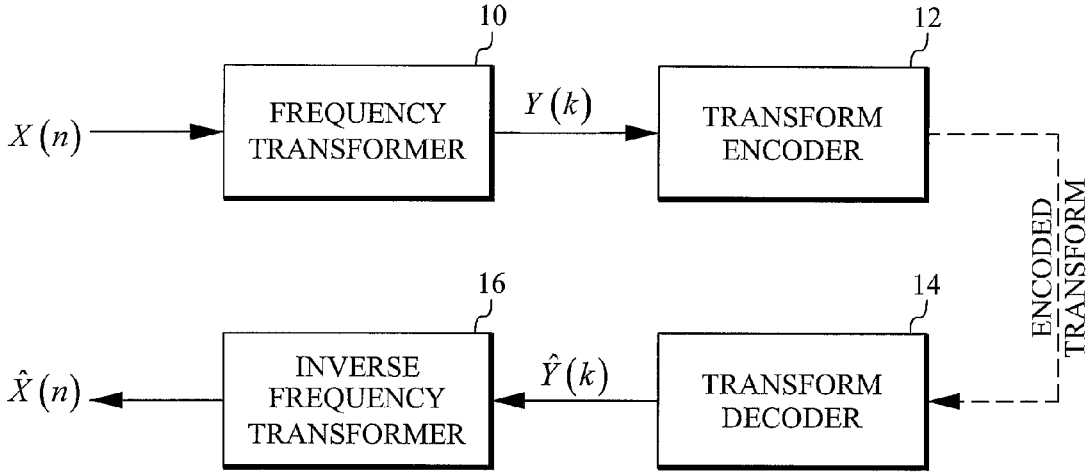

10
12
$X(n)$
FREQUENCY TRANSFORMER
$Y(k)$
TRANSFORM ENCODER
ENCODED TRANSFORM
16
14
$\hat{X}(n)$
INVERSE FREQUENCY TRANSFORMER
$\hat{Y}(k)$
TRANSFORM DECODER

FIG. 1

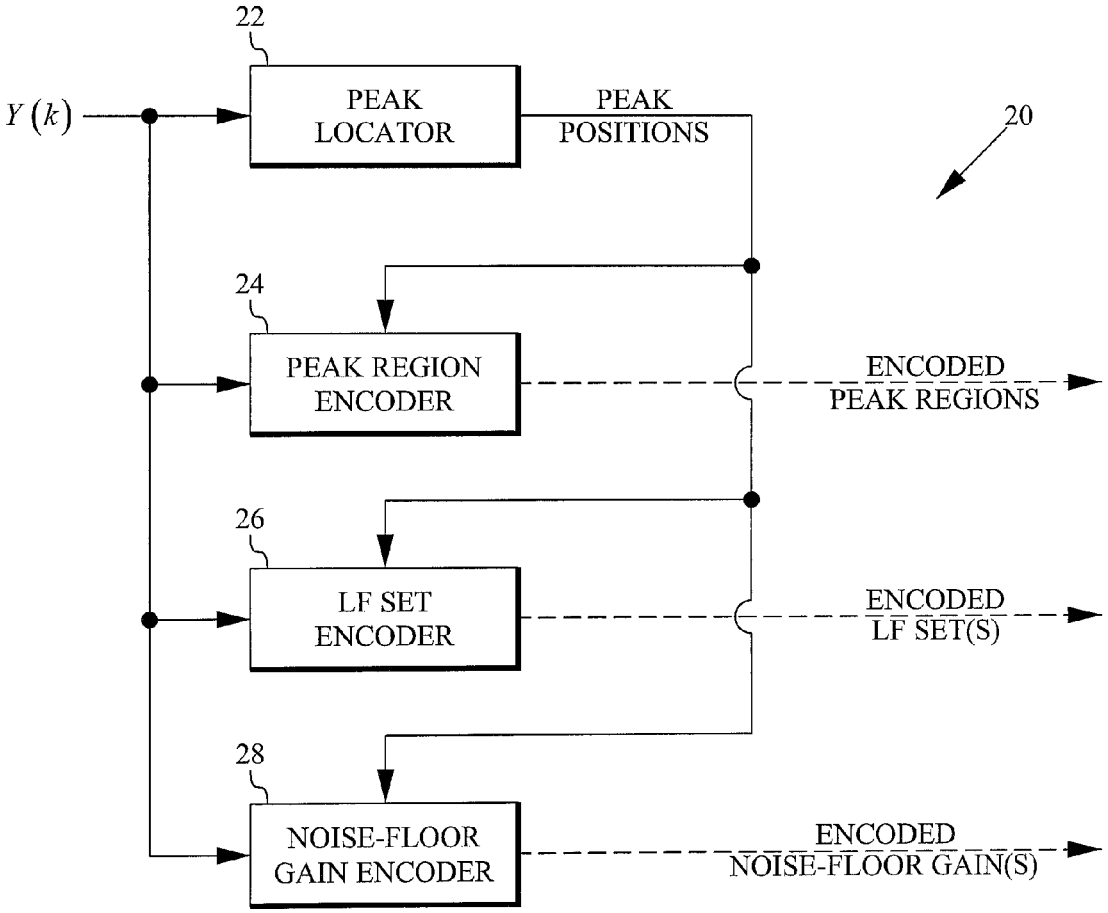

22
$Y(k)$
PEAK LOCATOR
PEAK POSITIONS
20
24
PEAK REGION ENCODER
ENCODED PEAK REGIONS
26
LF SET ENCODER
ENCODED LF SET(S)
28
NOISE-FLOOR GAIN ENCODER
ENCODED NOISE-FLOOR GAIN(S)

FIG. 7

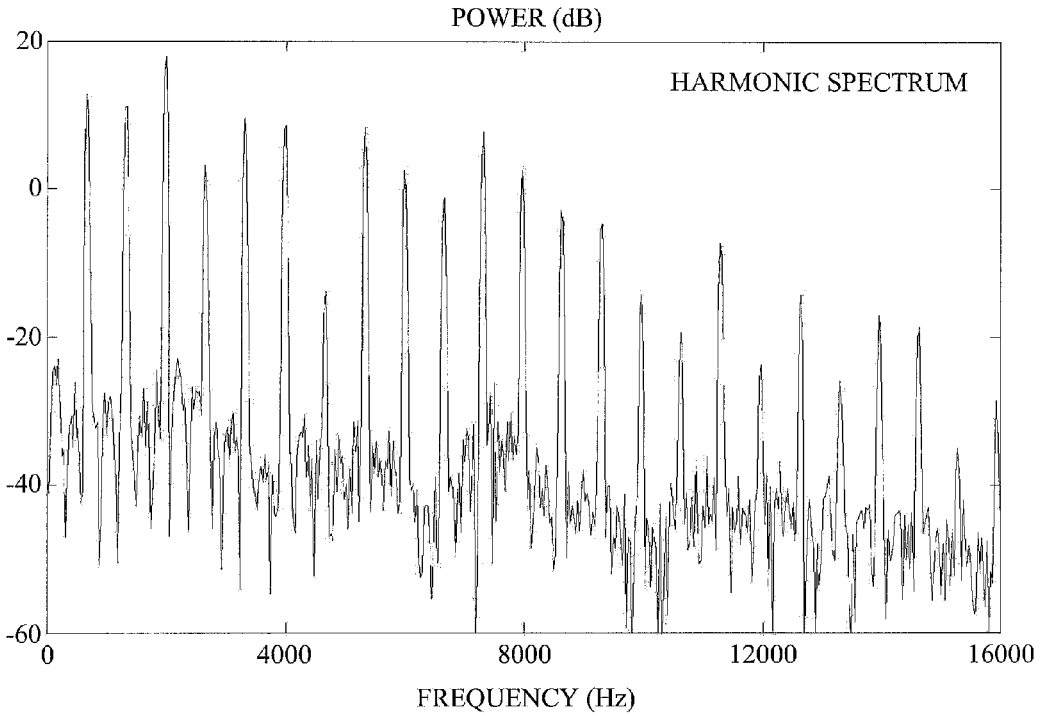

POWER (dB)
HARMONIC SPECTRUM
20
0
-20
-40
-60
0
4000
8000
12000
16000
FREQUENCY (Hz)

Fig. 2

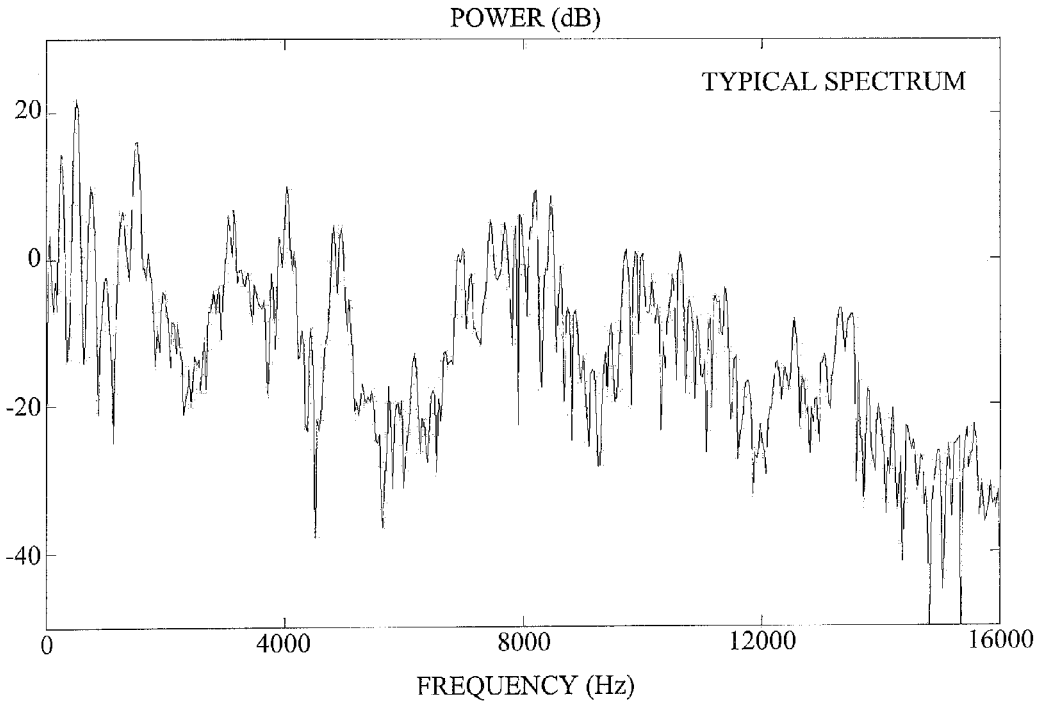

POWER (dB)
TYPICAL SPECTRUM
20
0
-20
-40
0
4000
8000
12000
16000
FREQUENCY (Hz)

Fig. 3

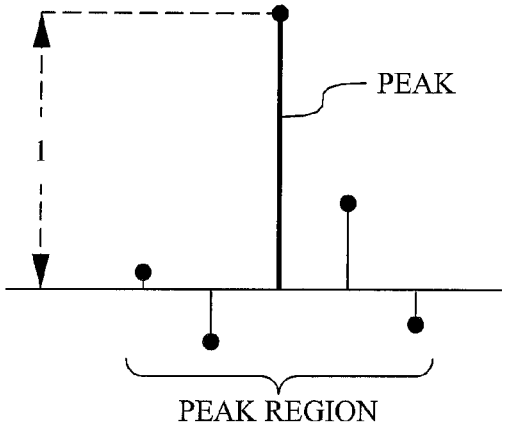

PEAK
1
PEAK REGION

FIG. 4

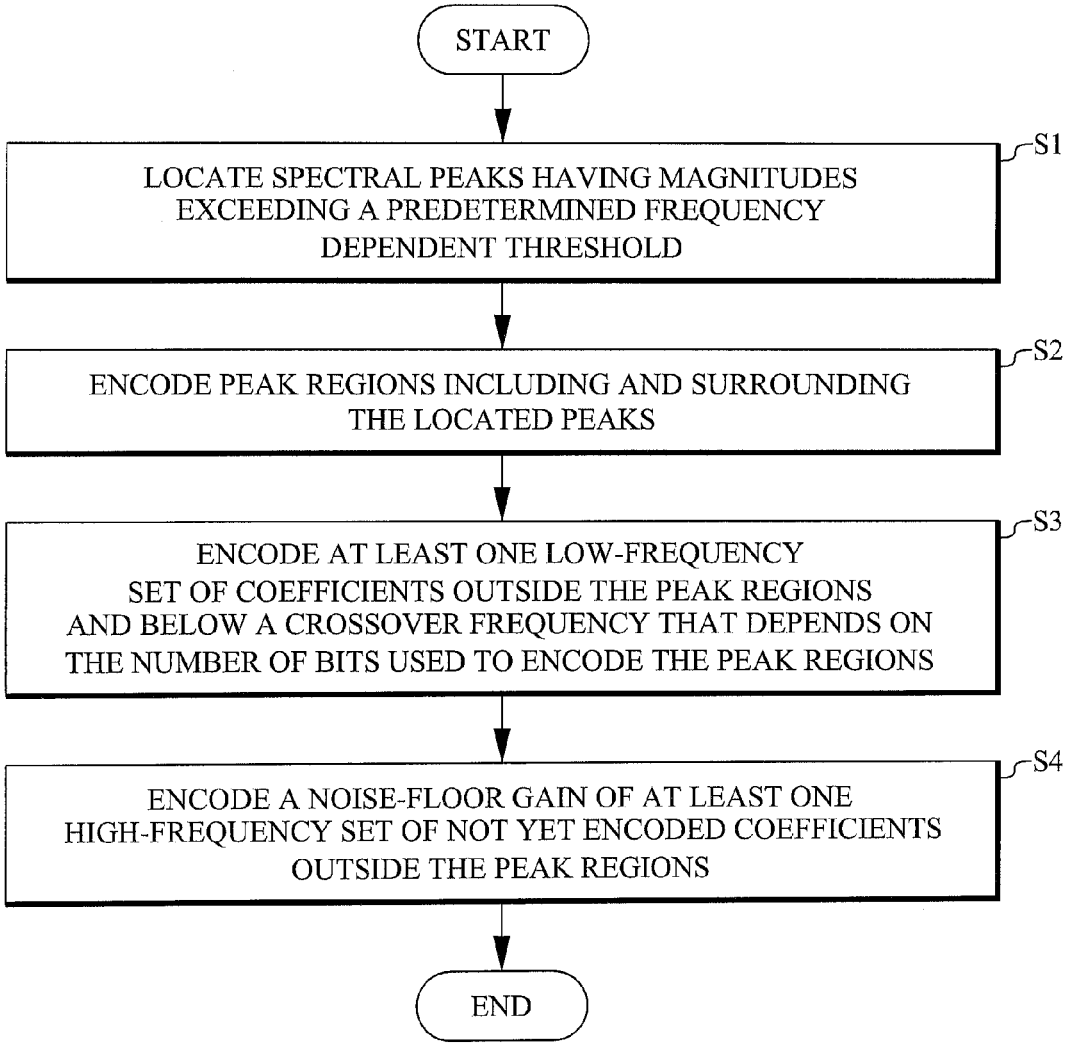

START
LOCATE SPECTRAL PEAKS HAVING MAGNITUDES EXCEEDING A PREDETERMINED FREQUENCY DEPENDENT THRESHOLD
S1
ENCODE PEAK REGIONS INCLUDING AND SURROUNDING THE LOCATED PEAKS
S2
ENCODE AT LEAST ONE LOW-FREQUENCY SET OF COEFFICIENTS OUTSIDE THE PEAK REGIONS AND BELOW A CROSSOVER FREQUENCY THAT DEPENDS ON THE NUMBER OF BITS USED TO ENCODE THE PEAK REGIONS
S3
ENCODE A NOISE-FLOOR GAIN OF AT LEAST ONE HIGH-FREQUENCY SET OF NOT YET ENCODED COEFFICIENTS OUTSIDE THE PEAK REGIONS
S4
END

FIG. 5

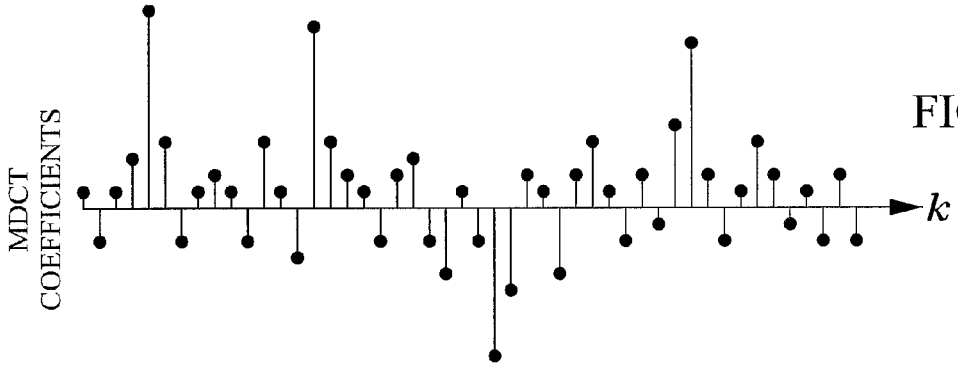


FIG. 6A

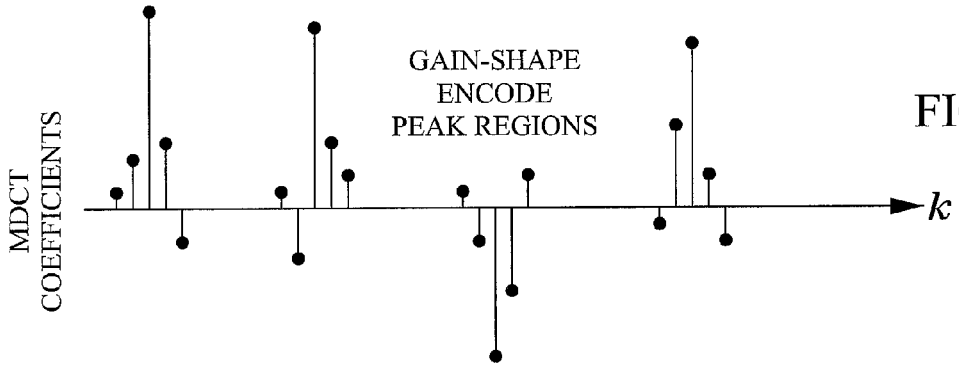


FIG. 6B

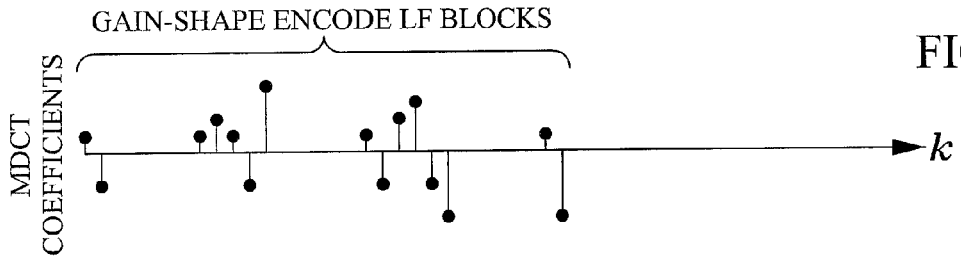


FIG. 6C

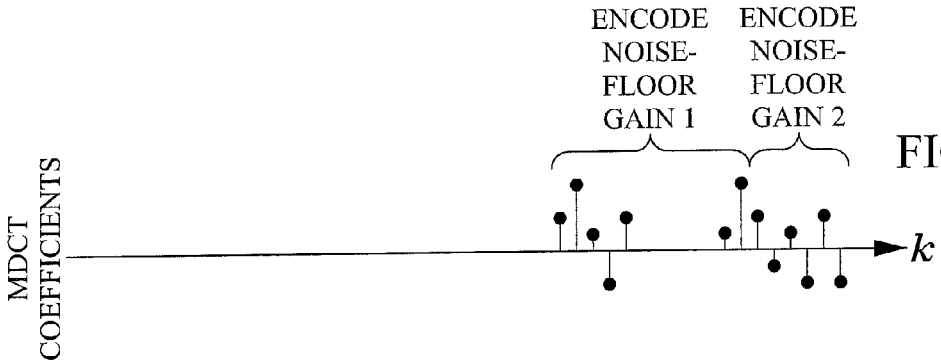


FIG. 6D

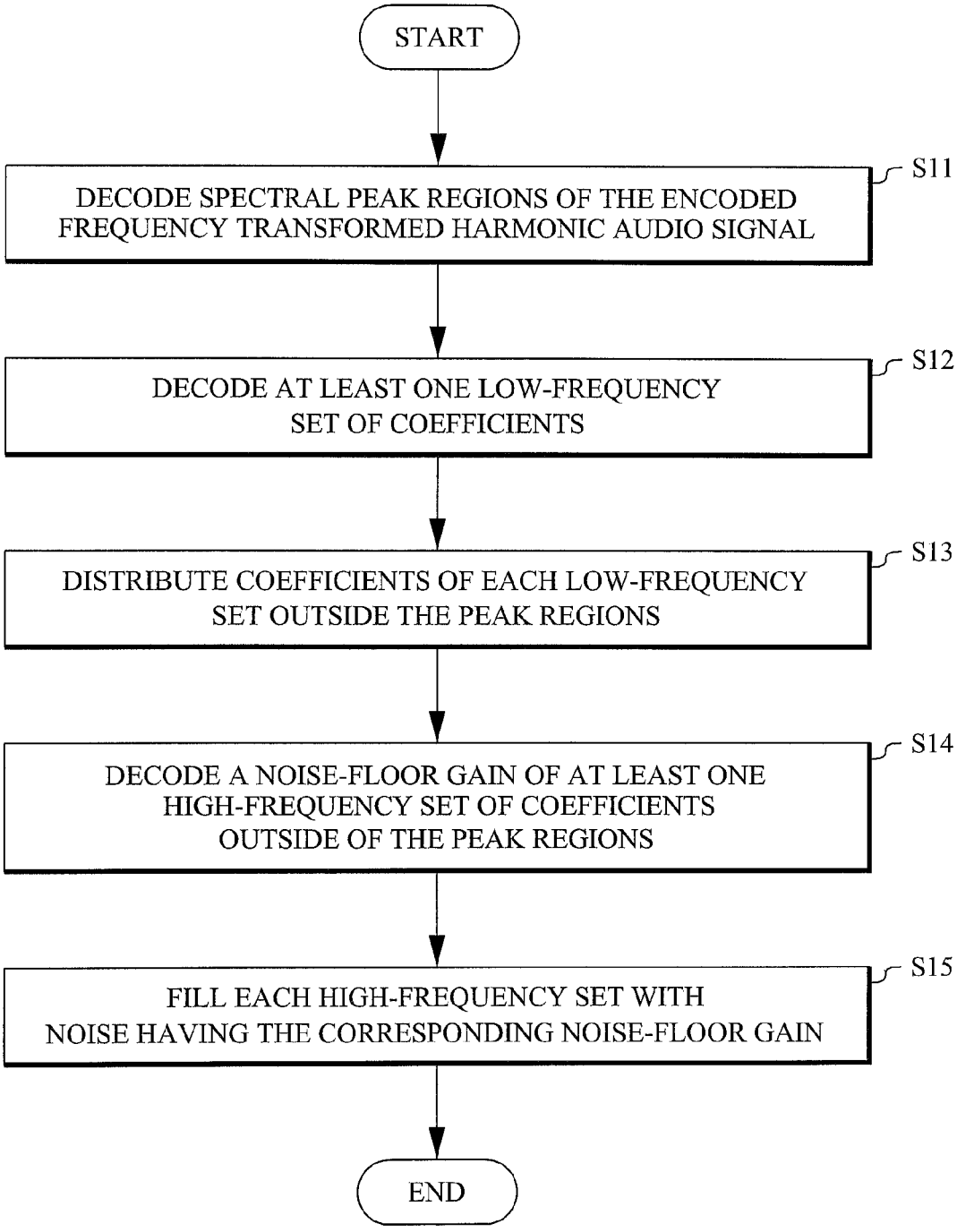

START
DECODE SPECTRAL PEAK REGIONS OF THE ENCODED FREQUENCY TRANSFORMED HARMONIC AUDIO SIGNAL
S11
DECODE AT LEAST ONE LOW-FREQUENCY SET OF COEFFICIENTS
S12
DISTRIBUTE COEFFICIENTS OF EACH LOW-FREQUENCY SET OUTSIDE THE PEAK REGIONS
S13
DECODE A NOISE-FLOOR GAIN OF AT LEAST ONE HIGH-FREQUENCY SET OF COEFFICIENTS OUTSIDE OF THE PEAK REGIONS
S14
FILL EACH HIGH-FREQUENCY SET WITH NOISE HAVING THE CORRESPONDING NOISE-FLOOR GAIN
S15
END

FIG. 8

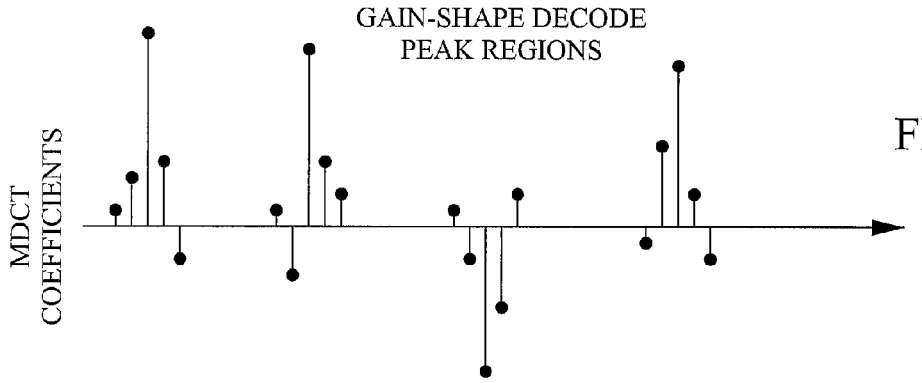


FIG. 9A

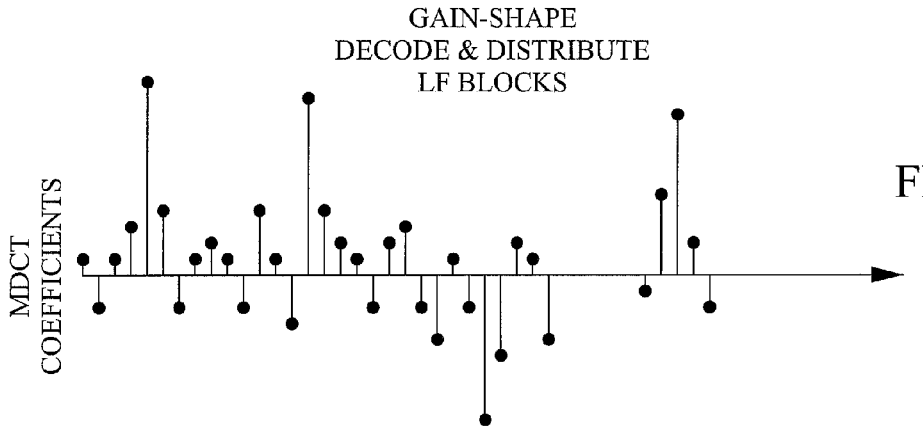


FIG. 9B

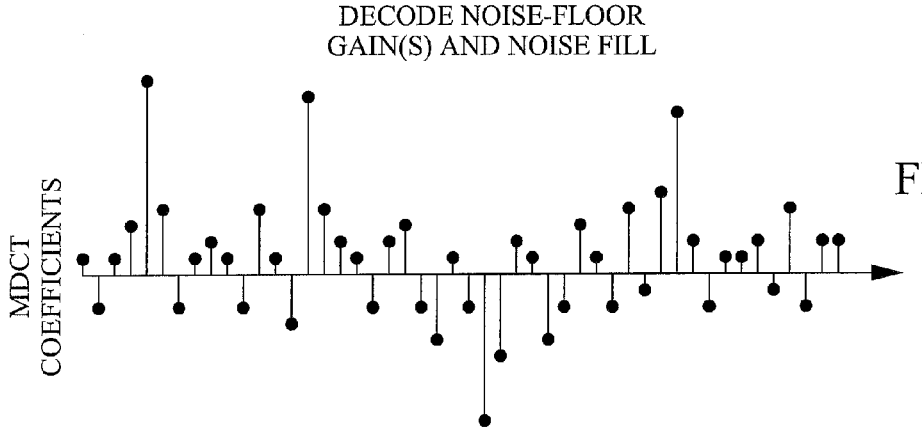


FIG. 9C

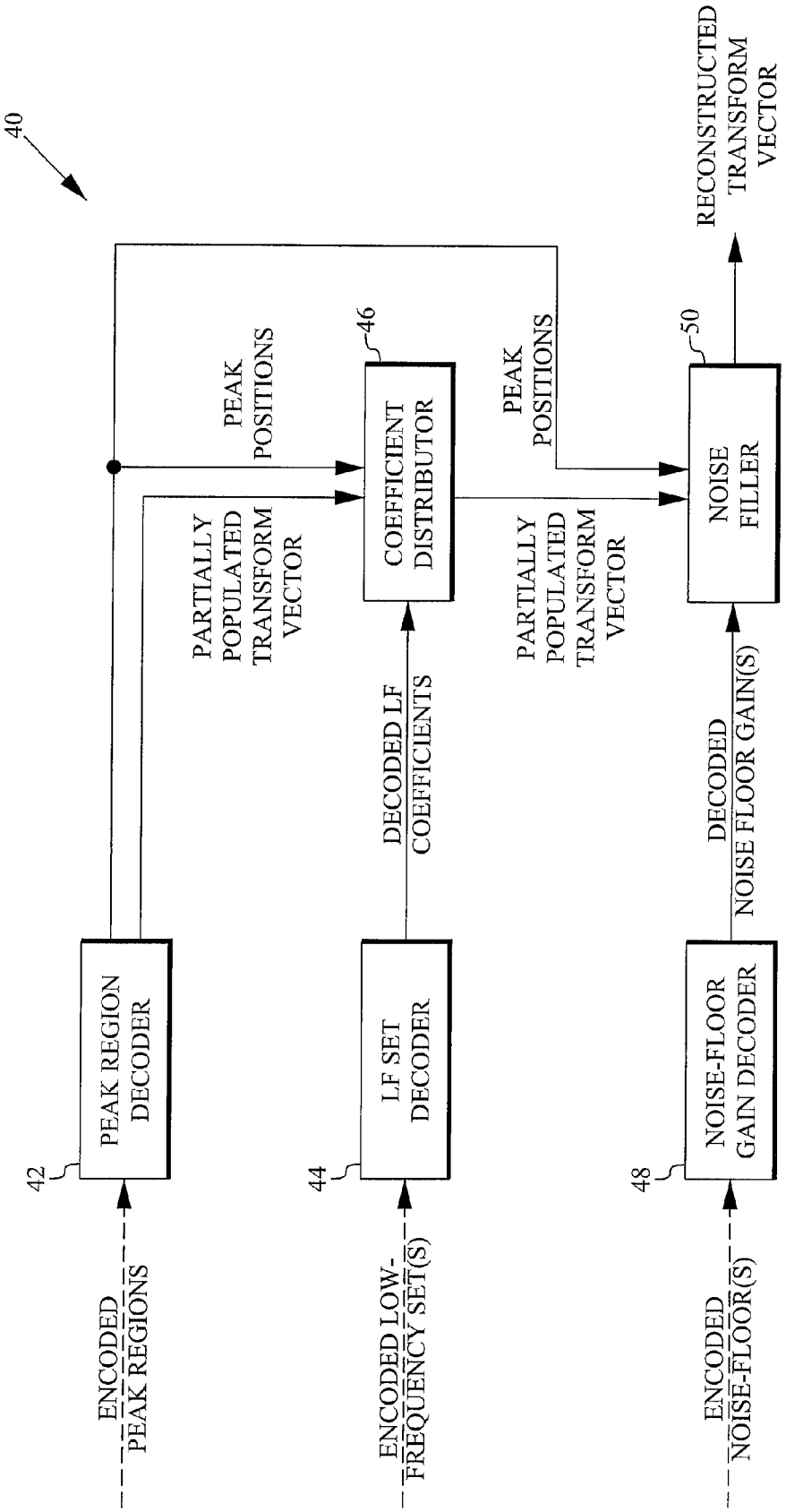

40
42
PEAK REGION DECODER
ENCODED PEAK REGIONS
44
LF SET DECODER
ENCODED LOW-FREQUENCY SET(S)
48
NOISE-FLOOR GAIN DECODER
ENCODED NOISE-FLOOR(S)
PARTIALLY POPULATED TRANSFORM VECTOR
PEAK POSITIONS
DECODED LF COEFFICIENTS
46
COEFFICIENT DISTRIBUTOR
PARTIALLY POPULATED TRANSFORM VECTOR
PEAK POSITIONS
DECODED NOISE FLOOR GAIN(S)
50
NOISE FILLER
RECONSTRUCTED TRANSFORM VECTOR

FIG. 10

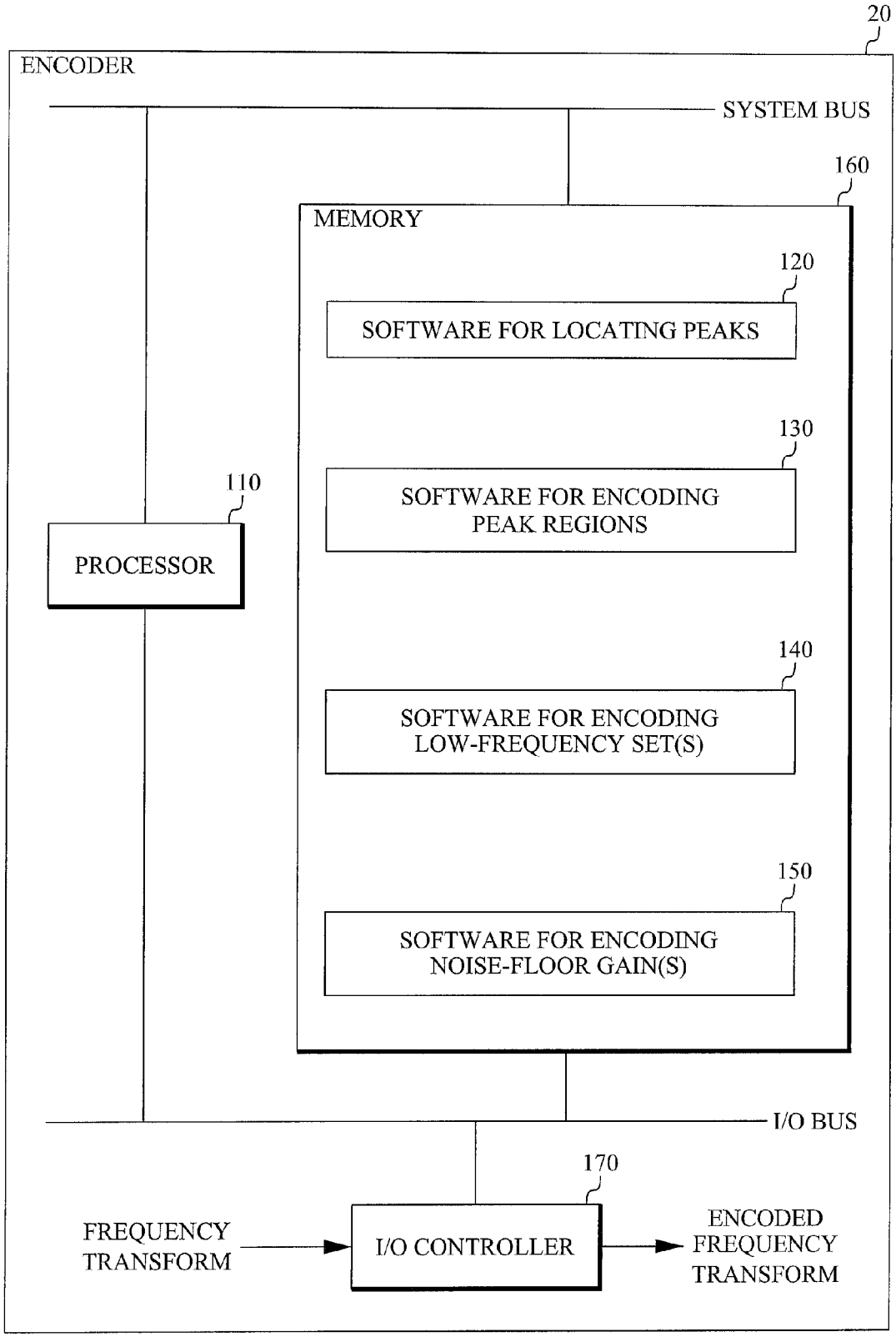

20
ENCODER
SYSTEM BUS
160
MEMORY
120
SOFTWARE FOR LOCATING PEAKS
130
SOFTWARE FOR ENCODING PEAK REGIONS
110
PROCESSOR
140
SOFTWARE FOR ENCODING LOW-FREQUENCY SET(S)
150
SOFTWARE FOR ENCODING NOISE-FLOOR GAIN(S)
I/O BUS
170
FREQUENCY TRANSFORM
I/O CONTROLLER
ENCODED FREQUENCY TRANSFORM

FIG. 11

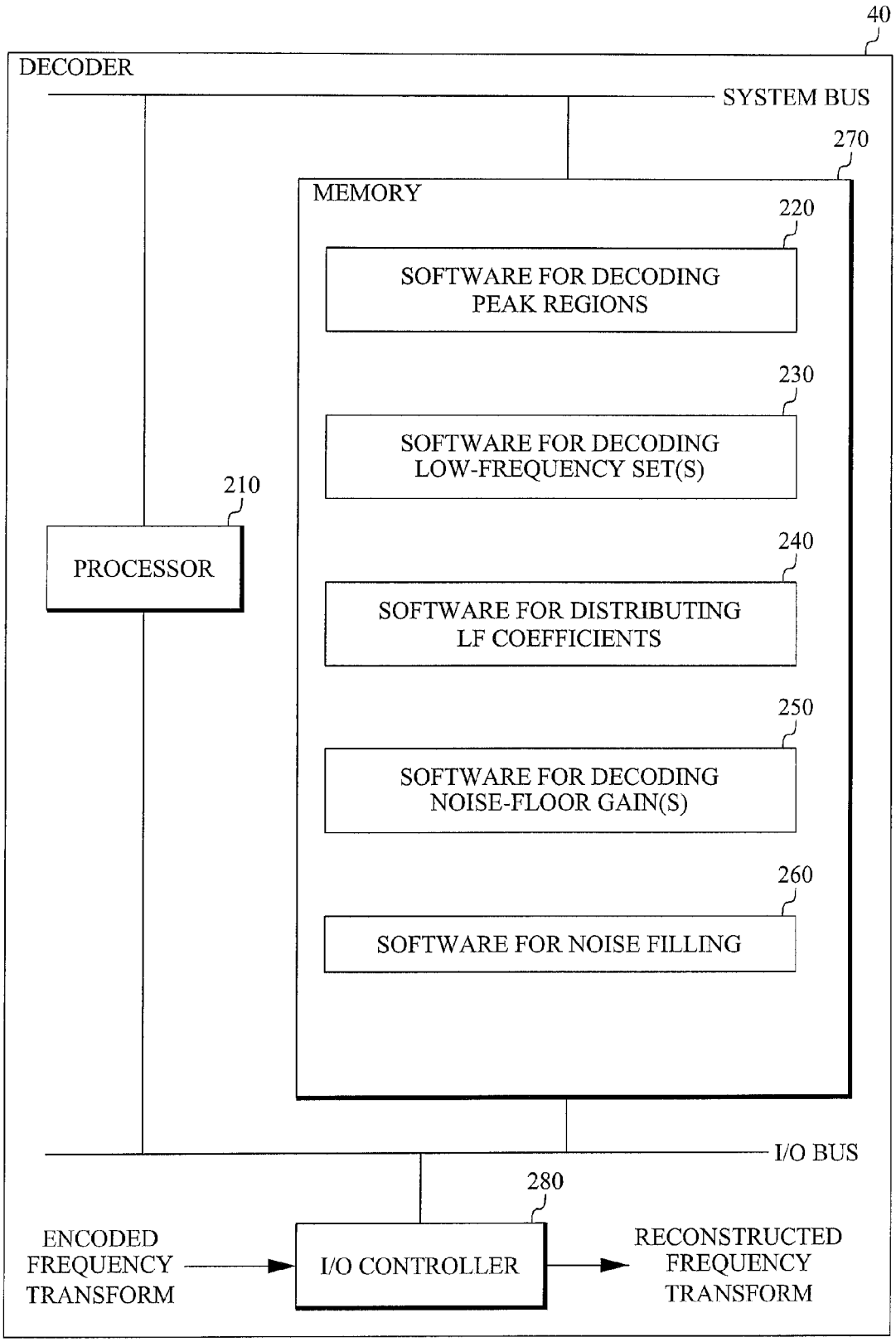

40
DECODER
SYSTEM BUS
270
MEMORY
220
SOFTWARE FOR DECODING PEAK REGIONS
230
SOFTWARE FOR DECODING LOW-FREQUENCY SET(S)
210
PROCESSOR
240
SOFTWARE FOR DISTRIBUTING LF COEFFICIENTS
250
SOFTWARE FOR DECODING NOISE-FLOOR GAIN(S)
260
SOFTWARE FOR NOISE FILLING
I/O BUS
280
ENCODED FREQUENCY TRANSFORM
I/O CONTROLLER
RECONSTRUCTED FREQUENCY TRANSFORM

FIG. 12

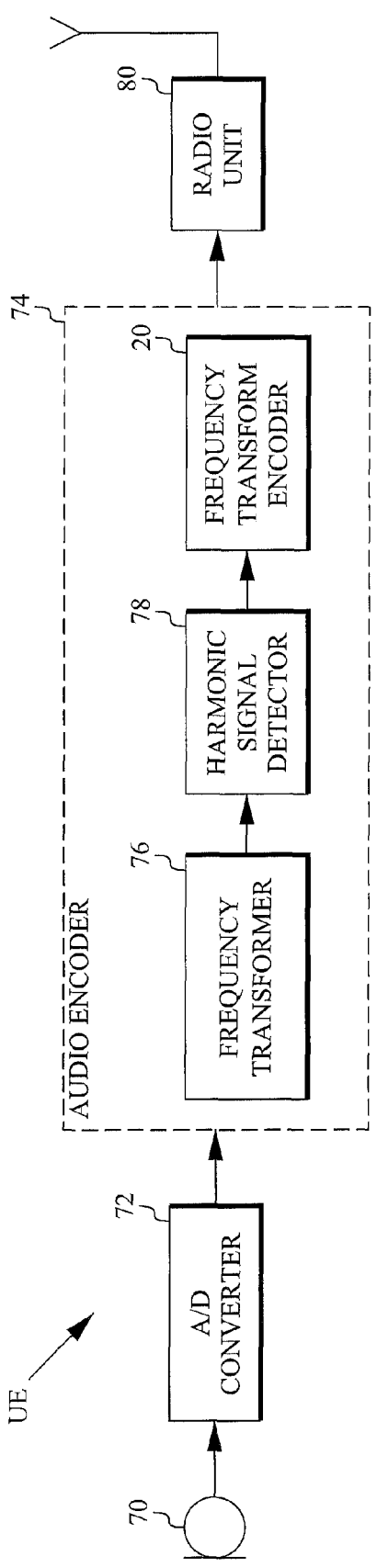

UE
70
A/D CONVERTER
72
AUDIO ENCODER
74
FREQUENCY TRANSFORMER
76
HARMONIC SIGNAL DETECTOR
78
FREQUENCY TRANSFORM ENCODER
20
RADIO UNIT
80

FIG. 13

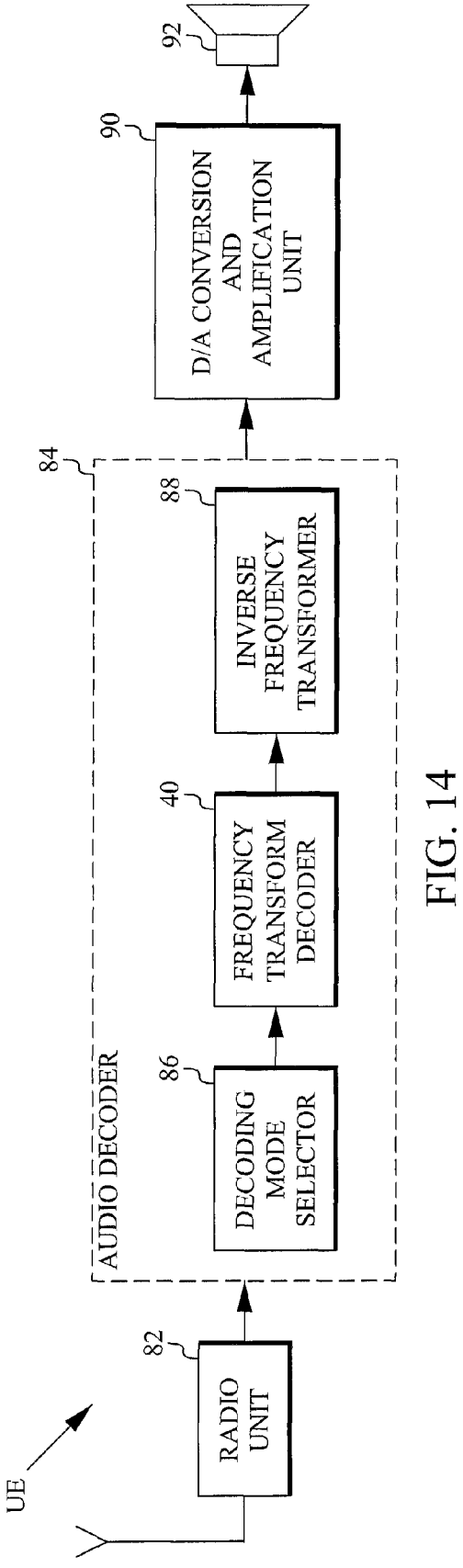

UE
RADIO UNIT
82
AUDIO DECODER
84
DECODING MODE SELECTOR
86
FREQUENCY TRANSFORM DECODER
40
INVERSE FREQUENCY TRANSFORMER
88
D/A CONVERSION AND AMPLIFICATION UNIT
90
92

FIG. 14

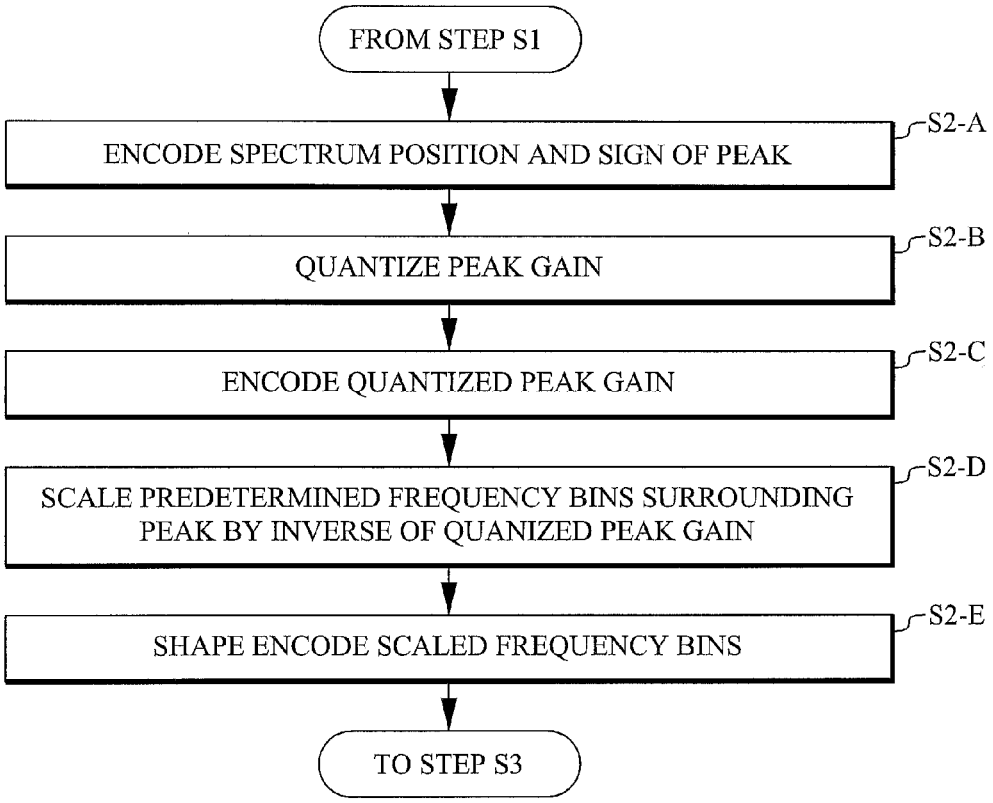

FROM STEP S1
ENCODE SPECTRUM POSITION AND SIGN OF PEAK
S2-A
QUANTIZE PEAK GAIN
S2-B
ENCODE QUANTIZED PEAK GAIN
S2-C
SCALE PREDETERMINED FREQUENCY BINS SURROUNDING PEAK BY INVERSE OF QUANIZED PEAK GAIN
S2-D
SHAPE ENCODE SCALED FREQUENCY BINS
S2-E
TO STEP S3

FIG. 15

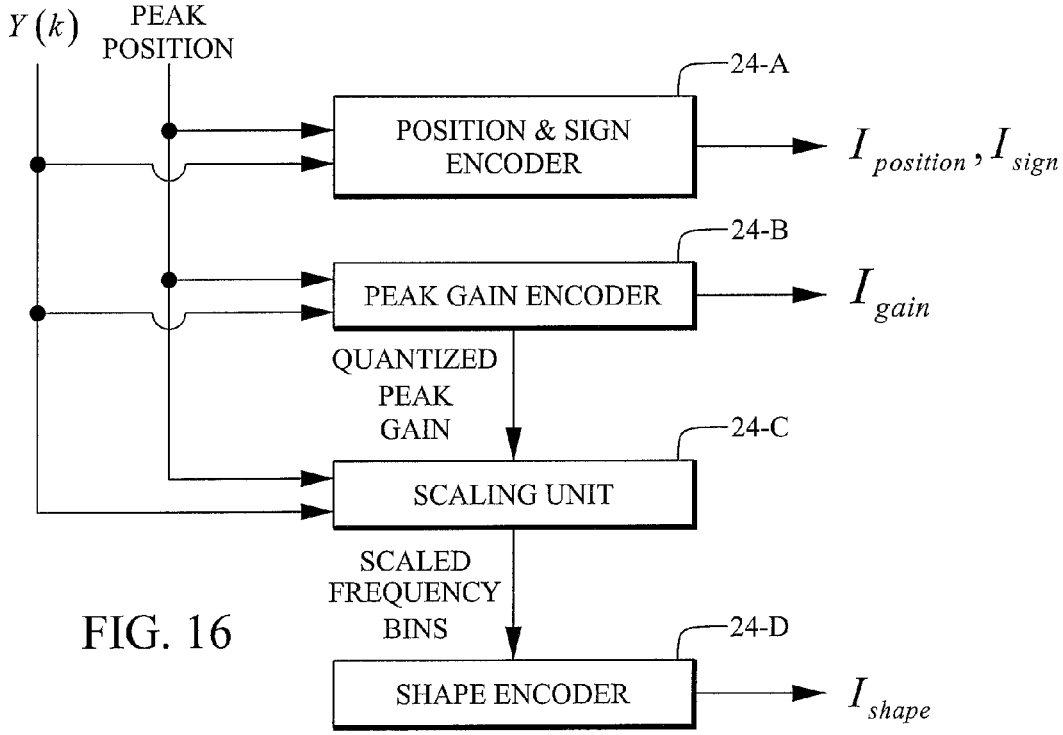

Y (k)
PEAK POSITION
24-A
POSITION & SIGN ENCODER
I position , I sign
24-B
PEAK GAIN ENCODER
I gain
QUANTIZED PEAK GAIN
24-C
SCALING UNIT
SCALED FREQUENCY BINS
24-D
SHAPE ENCODER
I shape

FIG. 16

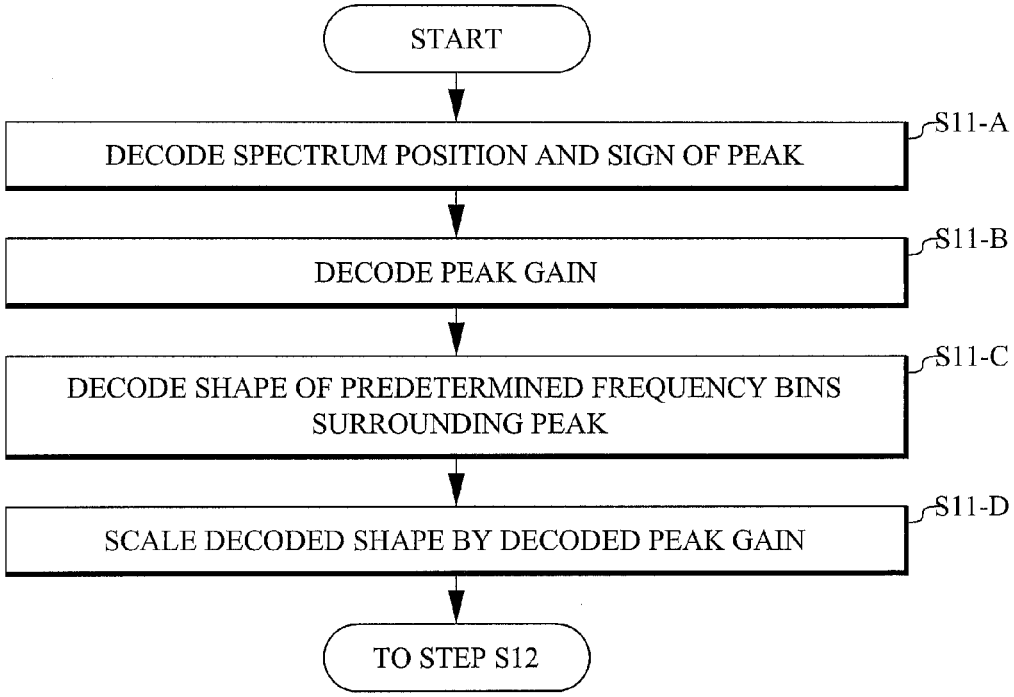

START
DECODE SPECTRUM POSITION AND SIGN OF PEAK
S11-A
DECODE PEAK GAIN
S11-B
DECODE SHAPE OF PREDETERMINED FREQUENCY BINS SURROUNDING PEAK
S11-C
SCALE DECODED SHAPE BY DECODED PEAK GAIN
S11-D
TO STEP S12

FIG. 17

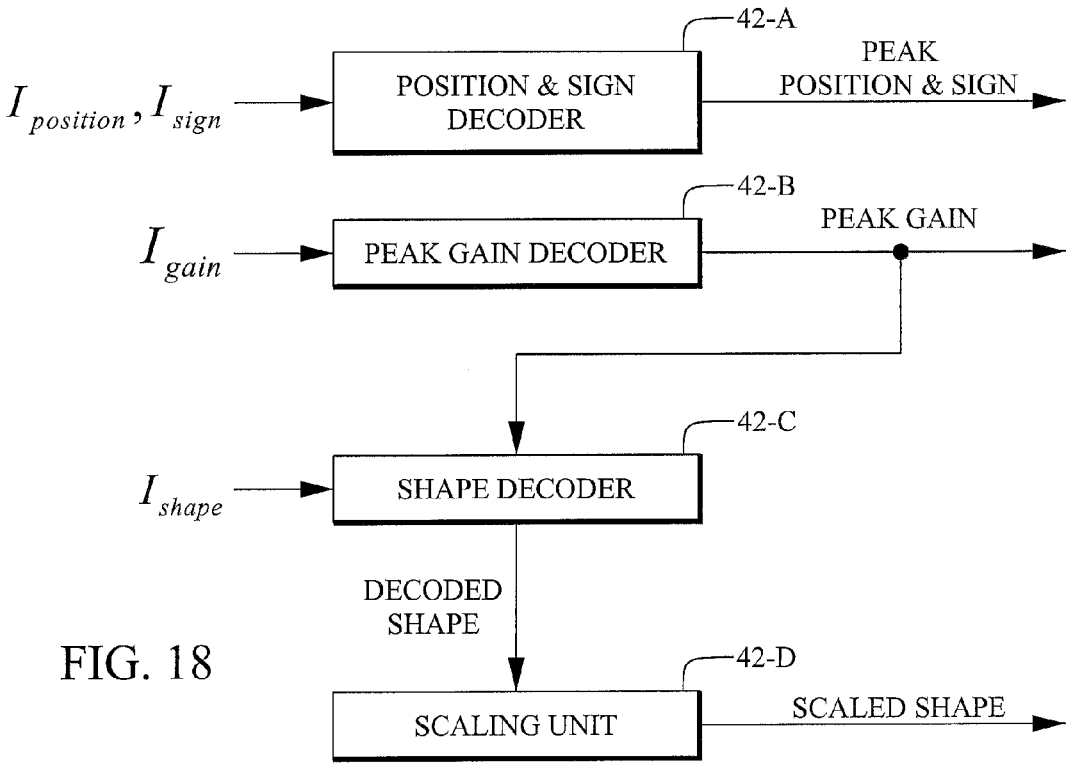

$I_{position}$, I_{sign}
POSITION & SIGN DECODER
42-A
PEAK POSITION & SIGN
I_{gain}
PEAK GAIN DECODER
42-B
PEAK GAIN
I_{shape}
SHAPE DECODER
42-C
DECODED SHAPE
SCALING UNIT
42-D
SCALED SHAPE

FIG. 18

TRANSFORM ENCODING/DECODING OF HARMONIC AUDIO SIGNALS

TECHNICAL FIELD

The proposed technology relates to transform encoding/decoding of audio signals, especially harmonic audio signals.

BACKGROUND

Transform encoding is the main technology used to compress and transmit audio signals. The concept of transform encoding is to first convert a signal to the frequency domain, and then to quantize and transmit the transform coefficients. The decoder uses the received transform coefficients to reconstruct the signal waveform by applying the inverse frequency transform, see FIG. **1**. In FIG. **1** an audio signal X(n) is forwarded to a frequency transformer **10**. The resulting frequency transform Y(k) is forwarded to a transform encoder **12**, and the encoded transform is transmitted to the decoder, where it is decoded by a transform decoder **14**. The decoded transform $\hat{Y}(k)$ is forwarded to an inverse frequency transformer **16** that transforms it into a decoded audio signal $\hat{X}(n)$. The motivation behind this scheme is that frequency domain coefficients can be more efficiently quantized for the following reasons:

1) Transform coefficients (Y(k) in FIG. **1**) are more uncorrelated than input signal samples (X(n) in FIG. **1**).
2) The frequency transform provides energy compaction (more coefficients Y(k) are close to zero and can be neglected), and
3) The subjective motivation behind the transform is that the human auditory system operates on a transformed domain, and it is easier to select perceptually important signal components on that domain.

In a typical transform codec the signal waveform is transformed on a block by block basis (with 50% overlap), using the Modified Discrete Cosine Transform (MDCT). In an MDCT type transform codec a block signal waveform X(n) is transformed into an MDCT vector Y(k). The length of the waveform blocks corresponds to 20-40 ms audio segments. If the length is denoted by 2L, the MDCT transform can be defined as:

$$Y(k) = \sqrt{\frac{2}{L}} \sum_{n=0}^{2L-1} \sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{L}\right] \cos\left[\left(n+\frac{1}{2}+\frac{1}{L}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{L}\right] X(n) \quad (1)$$

for k=0, . . . , L−1. Then the MDCT vector Y(k) is split into multiple bands (sub-vectors), and the energy (or gain) G(j) in each band is calculated as:

$$G(j) = \sqrt{\frac{1}{N_j} \sum_{k=m_j}^{m_j+N_j-1} Y^2(k)} \quad (2)$$

where m_j is the first coefficient in band j and N_1 refers to the number of MDCT coefficients in the corresponding bands (a typical range contains 8-32 coefficients). As an example of a uniform band structure, let N_j=8 for all j, then G(0) would be the energy of the first 8 coefficients, G(1) would be the energy of the next 8 coefficients, etc.

These energy values or gains give an approximation of the spectrum envelope, which is quantized, and the quantization indices are transmitted to the decoder. Residual sub-vectors or shapes are obtained by scaling the MDCT sub-vectors with the corresponding envelope gains, e.g. the residual in each band is scaled to have unit Root Mean Square (RMS) energy. Then the residual sub-vectors or shapes are quantized with different number of bits based on the corresponding envelope gains. Finally, at the decoder, the MDCT vector is reconstructed by scaling up the residual sub-vectors or shapes with the corresponding envelope gains, and an inverse MDCT is used to reconstruct the time-domain audio frame.

The conventional transform encoding concept does not work well with very harmonic audio signals, e.g. single instruments. An example of such a harmonic spectrum is illustrated in FIG. **2** (for comparison a typical audio spectrum without excessive harmonics is shown FIG. **3**). The reason is that the normalization with the spectrum envelope does not result in a sufficiently "flat" residual vector, and the residual encoding scheme cannot produce an audio signal of acceptable quality. This mismatch between the signal and the encoding model can be resolved only at very high bitrates, but in most cases this solution is not suitable.

SUMMARY

An object of the proposed technology is a transform encoding/decoding scheme that is more suited for harmonic audio signals.

The proposed technology involves a method of encoding frequency transform coefficients of a harmonic audio signal. The method includes the steps of:

- locating spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold;
- encoding peak regions including and surrounding the located peaks;
- encoding at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions;
- encoding a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

The proposed technology also involves an encoder for encoding frequency transform coefficients of a harmonic audio signal. The encoder includes:

- a peak locator configured to locate spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold;
- a peak region encoder configured to encode peak regions including and surrounding the located peaks;
- a low-frequency set encoder configured to encode at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions;
- a noise-floor gain encoder configured to encode a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

The proposed technology also involves a user equipment (UE) including such an encoder.

The proposed technology also involves a method of reconstructing frequency transform coefficients of an encoded frequency transformed harmonic audio signal. The method includes the steps of:

- decoding spectral peak regions of the encoded frequency transformed harmonic audio signal;

decoding at least one low-frequency set of coefficients;
distributing coefficients of each low-frequency set outside the peak regions;
decoding a noise-floor gain of at least one high-frequency set of coefficients outside of the peak regions;
filling each high-frequency set with noise having the corresponding noise-floor gain.

The proposed technology also involves a decoder for reconstructing frequency transform coefficients of an encoded frequency transformed harmonic audio signal. The decoder includes:

a peak region decoder configured to decode spectral peak regions of the encoded frequency transformed harmonic audio signal;
a low-frequency set decoder configured to decode at least one low-frequency set of coefficients;
a coefficient distributor configured to distribute coefficients of each low-frequency set outside the peak regions;
a noise-floor gain decoder configured to decode a noise-floor gain of at least one high-frequency set of coefficients outside of the peak regions;
a noise filler configured to fill each high-frequency set with noise having the corresponding noise-floor gain.

The proposed technology also involves a user equipment (UE) including such a decoder.

The proposed harmonic audio coding encoding/decoding scheme provides better perceptual quality than the conventional coding schemes for a large class of harmonic audio signals.

BRIEF DESCRIPTION OF THE DRAWINGS

The present technology, together with further objects and advantages thereof, may best be understood by making reference to the following description taken together with the accompanying drawings, in which:

FIG. **1** illustrates the frequency transform coding concept;

FIG. **2** illustrates a typical spectrum of a harmonic audio signal;

FIG. **3** illustrates a typical spectrum of a non-harmonic audio signal;

FIG. **4** illustrates a peak region;

FIG. **5** is a flow chart illustrating the proposed encoding method;

FIG. **6**A-D illustrates an example embodiment of the proposed encoding method;

FIG. **7** is a block diagram of an example embodiment of the proposed encoder;

FIG. **8** is a flow chart illustrating the proposed decoding method;

FIG. **9**A-C illustrates an example embodiment of the proposed decoding method;

FIG. **10** is a block diagram of an example embodiment of the proposed decoder;

FIG. **11** is a block diagram of an example embodiment of the proposed encoder;

FIG. **12** is a block diagram of an example embodiment of the proposed decoder;

FIG. **13** is a block diagram of an example embodiment of a UE including the proposed encoder;

FIG. **14** is a block diagram of an example embodiment of a UE including the proposed decoder;

FIG. **15** is a flow chart of an example embodiment of a part of the proposed encoding method;

FIG. **16** is block diagram of an example embodiment of a peak region encoder in the proposed encoder;

FIG. **17** is a flow chart of an example embodiment of a part of the proposed decoding method;

FIG. **18** is block diagram of an example embodiment of a peak region decoder in the proposed decoder.

DETAILED DESCRIPTION

FIG. **2** illustrates a typical spectrum of a harmonic audio signal, and FIG. **3** illustrates a typical spectrum of a non-harmonic audio signal. The spectrum of the harmonic signal is formed by strong spectral peaks separated by much weaker frequency bands, while the spectrum of the non-harmonic audio signal is much smoother.

The proposed technology provides an alternative audio encoding model that handles harmonic audio signals better. The main concept is that the frequency transform vector, for example an MDCT vector, is not split into envelope and residual part, but instead spectral peaks are directly extracted and quantized, together with neighboring MDCT bins. At high frequencies, low energy coefficients outside the peaks neighborhoods are not coded, but noise-filled at the decoder. Here the signal model used in the conventional encoding, {spectrum envelope+residual} is replaced with a new model {spectral peaks+noise-floor}. At low frequencies, coefficients outside the peak neighborhoods are still coded, since they have an important perceptual role.

Encoder

Major steps on the encoder side are:

Locate and code spectral peak regions
Code low-frequency (LF) spectral coefficients. The size of coded region depends on the number of bits remaining after peak region coding.
Code noise-floor gains for spectral coefficients outside the peak regions

First the noise-floor is estimated, then the spectral peaks are extracted by a peak picking algorithm (the corresponding algorithms are described in more detail in APPENDIX I-II). Each peak and its surrounding 4 neighbors are normalized to unit energy at the peak position, see FIG. **4**. In other words, the entire region is scaled such that the peak has amplitude one. The peak position, gain (represents peak amplitude, magnitude) and sign are quantized. A Vector Quantizer (VQ) is applied to the MDCT bins surrounding the peak and searches for the index I_{shape} of the codebook vector that provides the best match. The peak position, gain and sign, as well as the surrounding shape vectors are quantized and the quantization indices $\{I_{position}\ I_{gain}\ I_{sign}\ I_{shape}\}$ are transmitted to the decoder. In addition to these indices the decoder is also informed of the total number of peaks.

In the above example each peak region includes 4 neighbors that symmetrically surround the peak. However it is also feasible to have both fewer and more neighbors surrounding the peak in either symmetrical or asymmetrical fashion.

After the peak regions have been quantized, all available remaining bits (except reserved bits for noise-floor coding, see below) are used to quantize the low frequency MDCT coefficients. This is done by grouping the remaining unquantized MDCT coefficients into, for example, 24-dimensional bands starting from the first bin. Thus, these bands will cover the lowest frequencies up to a certain crossover frequency. Coefficients that have already been quantized in the peak coding are not included, so the bands are not necessarily made up from 24 consecutive coefficients. For this reason the bands will also be referred to as "sets" below.

The total number of LF bands or sets depends on the number of available bits, but there are always enough bits

reserved to create at least one set. When more bits are available the first set gets more bits assigned until a threshold for the maximum number of bits per set is reached. If there are more bits available another set is created and bits are assigned to this set until the threshold is reached. This procedure is repeated until all available bits have been spent. This means that the crossover frequency at which this process is stopped will be frame dependent, since the number of peaks will vary from frame to frame. The crossover frequency will be determined by the number of bits that are available for LF encoding once the peak regions have been encoded.

Quantization of the LF sets can be done with any suitable vector quantization scheme, but typically some type of gain-shape encoding is used. For example, factorial pulse coding may be used for the shape vector, and scalar quantizer may be used for the gain.

A certain number of bits are always reserved for encoding a noise-floor gain of at least one high-frequency band of coefficients outside the peak regions, and above the upper frequency of the LF bands. Preferably two gains are used for this purpose. These gains may be obtained from the noise-floor algorithm described in APPENDIX I. If factorial pulse coding is used for the encoding the low-frequency bands some LF coefficients may not be encoded. These coefficients can instead be included in the high-frequency band encoding. As in the case of the LF bands, the HF bands are not necessarily made up from consecutive coefficients. For this reason the bands will also be referred to as "sets" below.

If applicable, the spectrum envelope for a bandwidth extension (BWE) region is also encoded and transmitted. The number of bands (and the transition frequency where the BWE starts) is bitrate dependent, e.g. 5.6 kHz at 24 kbps and 6.4 kHz at 32 kbps.

FIG. **5** is a flow chart illustrating the proposed encoding method from a general perspective. Step S**1** locates spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold. Step S**2** encodes peak regions including and surrounding the located peaks. Step S**3** encodes at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions. Step S**4** encodes a noise-floor gain of at least one high-frequency set of not yet encoded (still uncoded or remaining) coefficients outside the peak regions.

FIG. **6**A-D illustrates an example embodiment of the proposed encoding method. FIG. **6**A illustrates the MDCT transform of the signal frame to be encoded. In the figure there are fewer coefficients than in an actual signal. However, it should be kept in mind that purpose of the figure is only to illustrate the encoding process. FIG. **6**B illustrates **4** identified peak regions ready for gain-shape encoding. The method described in APPENDIX II can be used to find them. Next the LF coefficients outside the peak regions are collected in FIG. **6**C. These are concatenated into blocks that are gain-shape encoded. The remaining coefficients of the original signal in FIG. **6**A are the high-frequency coefficients illustrated in FIG. **6**D. They are divided into 2 sets and encoded (as concatenated blocks) by a noise-floor gain for each set. This noise-floor gain can be obtained from the energy of each set or by estimates obtained from the noise-floor estimation algorithm described in APPENDIX I.

FIG. **7** is a block diagram of an example embodiment of a proposed encoder **20**. A peak locator **22** is configured to locate spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold. A peak region encoder **24** is configured to encode peak regions including and surrounding the extracted peaks. A low-frequency set encoder **26** is configured to encode at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions. A noise-floor gain encoder **28** is configured to encode a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions. In this embodiment the encoders **24**, **26**, **28** use the detected peak position to decide which coefficients to include in the respective encoding.

Decoder

Major steps on the decoder are:

Reconstruct spectral peak regions

Reconstruct LF spectral coefficients

Noise-fill non-coded regions with noise, scaled with the received noise-floor gains.

The audio decoder extracts, from the bit-stream, the number of peak regions and the quantization indices $\{I_{position}\ I_{gam}\ I_{sign}\ I_{shape}\}$ in order to reconstruct the coded peak regions. These quantization indices contain information about the spectral peak position, gain and sign of the peak, as well as the index for the codebook vector that provides the best match for the peak neighborhood.

The MDCT low-frequency coefficients outside the peak regions are reconstructed from the encoded LF coefficients.

The MDCT high-frequency coefficients outside the peak regions are noise-filled at the decoder. The noise-floor level is received by the decoder, preferably in the form of two coded noise-floor gains (one for the lower and one for the upper half or part of the vector).

If applicable, the audio decoder performs a BWE from a pre-defined transition frequency with the received envelope gains for HF MDCT coefficients.

FIG. **8** is a flow chart illustrating the proposed decoding method from a general perspective. Step S**11** decodes spectral peak regions of the encoded frequency transformed harmonic audio signal. Step S**12** decodes at least one low-frequency set of coefficients. Step S**13** distributes coefficients of each low-frequency set outside the peak regions. Step S**14** decodes a noise-floor gain of at least one high-frequency set of coefficients outside the peak regions. Step S**15** fills each high-frequency set with noise having the corresponding noise-floor gain.

In an example embodiment the decoding of a low-frequency set is based on a gain-shape decoding scheme.

In an example embodiment the gain-shape decoding scheme is based on scalar gain decoding and factorial pulse shape decoding.

An example embodiment includes the step of decoding a noise-floor gain for each of two high-frequency sets.

FIG. **9**A-C illustrates an example embodiment of the proposed decoding method. The reconstruction of the frequency transform starts by gain-shape decoding the spectral peak regions and their positions, as illustrated in FIG. **9**A. In FIG. **9**B the LF set(s) are gain-shape decoded and the decoded transform coefficient are distributed in blocks outside the peak regions. In FIG. **9**C the noise-floor gains are decoded and the remaining transform coefficients are filled with noise having corresponding noise-floor gains. In this way the transform of FIG. **6**A has been approximately reconstructed. A comparison of FIG. **9**C with FIGS. **6**A and **6**D shows that the noise filled regions have different individual coefficients but the same energy, as expected.

FIG. **10** is a block diagram of an example embodiment of a proposed decoder **40**. A peak region decoder **42** is configured to decode spectral peak regions of the encoded frequency transformed harmonic audio signal. A low-fre-

quency set decoder **44** is configured to decode at least one low-frequency set of coefficients. A coefficient distributor **46** configured to distribute coefficients of each low-frequency set outside the peak regions. A noise-floor gain decoder **48** is configured to decode a noise-floor of at least one high-frequency set of coefficients outside the peak regions. A noise filler **50** is configured to fill each high-frequency set with noise having the corresponding noise-floor gain. In this embodiment the peak positions are forwarded to the coefficient distributor **46** and the noise filler **50** to avoid overwriting of the peak regions.

The steps, functions, procedures and/or blocks described herein may be implemented in hardware using any conventional technology, such as discrete circuit or integrated circuit technology, including both general-purpose electronic circuitry and application-specific circuitry.

Alternatively, at least some of the steps, functions, procedures and/or blocks described herein may be implemented in software for execution by suitable processing equipment. This equipment may include, for example, one or several micro processors, one or several Digital Signal Processors (DSP), one or several Application Specific Integrated Circuits (ASIC), video accelerated hardware or one or several suitable programmable logic devices, such as Field Programmable Gate Arrays (FPGA). Combinations of such processing elements are also feasible.

It should also be understood that it may be possible to reuse the general processing capabilities already present in the encoder/decoder. This may, for example, be done by reprogramming of the existing software or by adding new software components.

FIG. **11** is a block diagram of an example embodiment of the proposed encoder **20**. This embodiment is based on a processor **110**, for example a micro processor, which executes software **120** for locating peaks, software **130** for encoding peak regions, software **140** for encoding at least one low-frequency set, and software **150** for encoding at least one noise-floor gain. The software is stored in memory **160**. The processor **110** communicates with the memory over a system bus. The incoming frequency transform is received by an input/output (I/O) controller **170** controlling an I/O bus, to which the processor **110** and the memory **160** are connected. The encoded frequency transform obtained from the software **150** is outputted from the memory **160** by the I/O controller **170** over the I/O bus.

FIG. **12** is a block diagram of an example embodiment of the proposed decoder **40**. This embodiment is based on a processor **210**, for example a micro processor, which executes software **220** for decoding peak regions, software **230** for decoding at least one low-frequency set, software **240** for distributing LF coefficients, software **250** for decoding at least one noise-floor gain, and software **260** for noise filling. The software is stored in memory **270**. The processor **210** communicates with the memory over a system bus. The incoming encoded frequency transform is received by an input/output (I/O) controller **280** controlling an I/O bus, to which the processor **210** and the memory **280** are connected. The reconstructed frequency transform obtained from the software **260** is outputted from the memory **270** by the I/O controller **280** over the I/O bus.

The technology described above is intended to be used in an audio encoder/decoder, which can be used in a mobile device (e.g. mobile phone, laptop) or a stationary device, such as a personal computer. Here the term User Equipment (UE) will be used as a generic name for such devices.

FIG. **13** is a block diagram of an example embodiment of a UE including the proposed encoder. An audio signal from a microphone **70** is forwarded to an A/D converter **72**, the output of which is forwarded to an audio encoder **74**. The audio encoder **74** includes a frequency transformer **76** transforming the digital audio samples into the frequency domain. A harmonic signal detector **78** determines whether the transform represents harmonic or non-harmonic audio. If it represents non-harmonic audio, it is encoded in a conventional encoding mode (not shown). If it represents harmonic audio, it is forwarded to a frequency transform encoder **20** in accordance with the proposed technology. The encoded signal is forwarded to a radio unit **80** for transmission to a receiver.

The decision of the harmonic signal detector **78** is based on the noise-floor energy $\bar{E}_{nf}$ and peak energy $\bar{E}_p$ in APPENDIX I and II. The logic is as follows: IF $\bar{E}_p/\bar{E}_{nf}$ is above a threshold AND the number of detected peaks is in a predefined range THEN the signal is classified as harmonic. Otherwise the signal is classified as non-harmonic. The classification and thus the encoding mode is explicitly signaled to the decoder.

FIG. **14** is a block diagram of an example embodiment of a UE including the proposed decoder. A radio signal received by a radio unit **82** is converted to baseband, channel decoded and forwarded to an audio decoder **84**. The audio decoder includes a decoding mode selector **86**, which forwards the signal a frequency transform decoder **40** in accordance with the proposed technology if it has been classified as harmonic. If it has been classified as non-harmonic audio, it is decoded in a conventional decoder (not shown). The frequency transform decoder **40** reconstructs the frequency transform as described above. The reconstructed frequency transform is converted to the time domain in an inverse frequency transformer **88**. The resulting audio samples are forwarded to a D/A conversion and amplification unit **90**, which forwards the final audio signal to a loudspeaker **92**.

FIG. **15** is a flow chart of an example embodiment of a part of the proposed encoding method. In this embodiment the peak region encoding step S**2** in FIG. **5** has been divided into sub-steps S**2**-A to S**2**-E. Step S**2**-A encodes spectrum position and sign of a peak. Step S**2**-B quantizes peak gain. Step S**2**-C encodes the quantized peak gain. Step S**2**-D scales predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain. Step S**2**-E shape encodes the scaled frequency bins.

FIG. **16** is block diagram of an example embodiment of a peak region encoder in the proposed encoder. In this embodiment the peak region encoder **24** includes elements **24**-A to **24**-D. Position and sign encoder **24**-A is configured to encode spectrum position and sign of a peak. Peak gain encoder **24**-B is configured to quantize peak gain and to encode the quantized peak gain. Scaling unit **24**-C is configured to scale predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain. Shape encoder **24**-D is configured to shape encode the scaled frequency bins.

FIG. **17** is a flow chart of an example embodiment of a part of the proposed decoding method. In this embodiment the peak region decoding step S**11** in FIG. **8** has been divided into sub-steps S**11**-A to S**11**-D. Step S**11**-A decodes spectrum position and sign of a peak. Step S**11**-B decodes peak gain. Step S**11**-C decodes a shape of predetermined frequency bins surrounding the peak. Step S**11**-D scales the decoded shape by the decoded peak gain.

FIG. **18** is block diagram of an example embodiment of a peak region decoder in the proposed decoder. In this embodiment the peak region decoder **42** includes elements

42-A to 42-D. A position and sign decoder 42-A is configured to decode spectrum position and sign of a peak. A peak gain decoder 42-B is configured to decode peak gain. A shape decoder 42-C is configured to decode a shape of predetermined frequency bins surrounding the peak. A scaling unit 42-D is configured to scale the decoded shape by the decoded peak gain.

Specific implementation details for a 24 kbps mode are given below.

- The codec operates on 20 ms frames, which at a bit rate of 24 kbps gives 480 bits per-frame.
- The processed audio signal is sampled at 32 kHz, and has an audio bandwidth of 16 kHz.
- The transition frequency is set to 5.6 kHz (all frequency components above 5.6 kHz are bandwidth-extended).
- Reserved bits for signaling and bandwidth extension of frequencies above the transition frequency: ~30-40.
- Bits for coding two noise-floor gains: 10.
- The number of coded spectral peak regions is 7-17. The number of bits used per peak region is ~20-22, which gives a total number of ~140-340 for coding all peaks positions, gains, signs, and shapes.
- Bits for coding low frequency bands: ~100-300
- Coded low frequency bands: 1-4 (each band contains 8 MDCT bins). Since each MDCT bin corresponds to 25 Hz, coded low-frequency region corresponds to 200-800 Hz
- The gains used for bandwidth extension and the peak gains are Huffman coded so the number of bits used by these might vary between frames even for a constant number of peaks.
- The peak position and sign coding makes use of an optimization which makes it more efficient as the number of peaks increase. For 7 peaks, position and sign requires about 6.9 bits per peak and for 17 peaks the number is about 5.7 bits per peak.
- This variability in how many bits are used in different stages of the coding is no problem since the low frequency band coding comes last and just uses up whatever bits remain. However the system is designed so that enough bits always remain to encode one low frequency band.

The table below presents results from a listening test performed in accordance with the procedure described in ITU-R BS.1534-1 MUSHRA (Multiple Stimuli with Hidden Reference and Anchor). The scale in a MUSHRA test is 0 to 100, where low values correspond to low perceived quality, and high values correspond to high quality. Both codecs operated at 24 kbps. Test results are averaged over 24 music items and votes from 8 listeners.

System Under Test	MUSHRA Score
Low-pass anchor signal (bandwidth 7 kHz)	48.89
Conventional coding scheme	49.94
Proposed harmonic coding scheme	55.87
Reference signal (bandwidth 16 kHz)	100.00

It will be understood by those skilled in the art that various modifications and changes may be made to the proposed technology without departure from the scope thereof, which is defined by the appended claims.

Appendix I

The noise-floor estimation algorithm operates on the absolute values of transform coefficients |Y(k)|. Instantaneous noise-floor energies $E_{nf}(k)$ are estimated according to the recursion:

$$E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)| \quad (3)$$

where

$$\alpha = \begin{cases} 0.9578 \text{ if } |Y(k)| > E_{nf}(k-1) \\ 0.6472 \text{ if } |Y(k)| \le E_{nf}(k-1) \end{cases} \quad (4)$$

The particular form of the weighting factor α minimizes the effect of high-energy transform coefficients and emphasizes the contribution of low-energy coefficients. Finally the noise-floor level $\bar{E}_{nf}$ is estimated by simply averaging the instantaneous energies $E_{nf}(k)$.

Appendix II

The peak-picking algorithm requires knowledge of noise-floor level and average level of spectral peaks. The peak energy estimation algorithm is similar to the noise-floor estimation algorithm, but instead of low-energy, it tracks high-spectral energies:

$$E_p(k) = \beta E_p(k) + (1-\beta)|Y(k)| \quad (5)$$

where

$$\beta = \begin{cases} 0.4223 \text{ if } |Y(k)| > E_p(k-1) \\ 0.8029 \text{ if } |Y(k)| \le E_p(k-1) \end{cases} \quad (6)$$

In this case the weighting factor β minimizes the effect of low-energy transform coefficients and emphasizes the contribution of high-energy coefficients. The overall peak energy $\bar{E}_p$ is estimated by simply averaging the instantaneous energies.

When the peak and noise-floor levels are calculated, a threshold level θ is formed as:

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{E}_{nf} \quad (7)$$

with γ=0.88579. Transform coefficients are compared to the threshold, and the ones with amplitude above it, form a vector of peak candidates. Since the natural sources do not typically produce peaks that are very close, e.g., 80 Hz, the vector with peak candidates is further refined. Vector elements are extracted in decreasing order, and the neighborhood of each element is set to zero. In this way only the largest element in certain spectral region remain, and the set of these elements form the spectral peaks for the current frame.

Abbreviations

ASIC Application Specific Integrated Circuit
BWE BandWidth Extension
DSP Digital Signal Processors
FPGA Field Programmable Gate Arrays
HF High-Frequency
LF Low-Frequency
MDCT Modified Discrete Cosine Transform
RMS Root Mean Square
VQ Vector Quantizer

The invention claimed is:

1. A method of encoding a frequency transformed harmonic audio signal, comprising:
receiving the frequency transformed harmonic audio signal;
generating an encoded frequency transformed harmonic audio signal corresponding to the frequency transformed harmonic audio signal, based on:
locating spectral peaks in the frequency transformed harmonic audio signal that have magnitudes exceeding a predetermined frequency dependent threshold;
encoding peak regions including and surrounding the located spectral peaks;
encoding at least one low-frequency set of Modified Discrete Cosine Transform (MDCT) coefficients outside the peak regions and below a crossover frequency that depends on a number of bits used to encode the peak regions;
encoding a noise-floor gain of at least one high-frequency set of not yet encoded MDCT coefficients outside the peak regions; and
outputting the encoded frequency transformed harmonic audio signal.

2. The encoding method of claim 1, wherein a peak region is encoded by:
encoding spectrum position and sign of a peak;
quantizing peak gain;
encoding the quantized peak gain;
scaling predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain; and
shape encoding the scaled frequency bins.

3. The encoding method of claim 1, wherein encoding a low-frequency set of MDCT coefficients includes encoding the low-frequency set based on a gain-shape encoding scheme.

4. The encoding method of claim 3, wherein the gain-shape encoding scheme is based on scalar gain quantization and factorial pulse shape encoding.

5. The encoding method of claim 1, comprising encoding a noise-floor gain for each of two high-frequency sets.

6. A method of audio signal reconstruction comprising:
receiving an encoded frequency transformed harmonic audio signal;
decoding the encoded frequency transformed harmonic audio signal and thereby obtaining a reconstructed frequency transformed harmonic audio signal, based on:
decoding spectral peak regions of the encoded frequency transformed harmonic audio signal, said spectral peak regions comprising spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold;
decoding at least one low-frequency set of Modified Discrete Cosine Transform (MDCT) coefficients of the encoded frequency transformed harmonic audio signal;
distributing the MDCT coefficients of each low-frequency set outside the spectral peak regions and below a crossover frequency that depends on a number of bits used to encode the peak regions;
decoding a noise-floor gain of at least one high-frequency set of MDCT coefficients of the encoded frequency transformed harmonic audio signal that are outside of the spectral peak regions;
filling each high-frequency set of MDCT coefficients with noise having the corresponding decoded noise-floor gain; and
outputting the reconstructed frequency transform harmonic audio signal.

7. The reconstruction method of claim 6, wherein a peak region is decoded by:
decoding spectrum position and sign of a peak;
decoding peak gain;
decoding a shape of predetermined frequency bins surrounding the peak; and
scaling the decoded shape by the decoded peak gain.

8. The reconstruction method of claim 6, wherein decoding a low-frequency set includes decoding the low-frequency set based on a gain-shape decoding scheme.

9. The reconstruction method of claim 8, wherein the gain-shape decoding scheme is based on scalar gain decoding and factorial pulse shape decoding.

10. The reconstruction method of claim 6, comprising decoding a noise-floor gain for each of two high-frequency sets.

11. An encoder for encoding a frequency transformed harmonic audio signal, said encoder configured to obtain the frequency transformed harmonic audio signal and comprising a processing circuit configured to:
generate an encoded frequency transformed harmonic audio signal corresponding to the frequency transformed harmonic audio signal, based on being configured to:
locate spectral peaks in the frequency transformed harmonic audio signal that have magnitudes exceeding a predetermined frequency dependent threshold;
encode peak regions including and surrounding the located spectral peaks;
encode at least one low-frequency set of Modified Discrete Cosine Transform (MDCT) coefficients outside the peak regions and below a crossover frequency that depends on a number of bits used to encode the peak regions; and
encode a noise-floor gain of at least one high-frequency set of not yet encoded MDCT coefficients outside the peak regions; and
output the encoded frequency transformed harmonic audio signal.

12. The encoder of claim 11, wherein the processing circuit is configured to:
encode a spectrum position and sign of a peak;
quantize peak gain and encode the quantized peak gain;
scale predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain; and
shape encode the scaled frequency bins.

13. A user equipment (UE) comprising the encoder of claim 11, said encoder configured to output the encoded frequency transformed harmonic audio signal to radio circuitry of the UE, for transmission to a remote receiver.

14. A decoder configured for audio signal reconstruction, said decoder configured to receive an encoded frequency transformed harmonic audio signal and comprising a processing circuit configured to:
decode the encoded frequency transformed harmonic audio signal and thereby obtain a reconstructed frequency transformed harmonic audio signal, based on being configured to:
decode spectral peak regions of the encoded frequency transformed harmonic audio signal, said spectral peak regions including spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold;
decode at least one low-frequency set of Modified Discrete Cosine Transform (MDCT) coefficients;

distribute the MDCT coefficients of each low-frequency set outside the spectral peak regions and below a crossover frequency that depends on a number of bits used to encode the peak regions;

decode a noise-floor gain of at least one high-frequency set of MDCT coefficients outside of the spectral peak regions; and

fill each high-frequency set of MDCT coefficients with noise having the corresponding noise-floor gain; and

output the reconstructed frequency transformed harmonic audio signal.

15. The decoder of claim **14**, wherein the processing circuit is configured to:

decode spectrum position and sign of a peak;

decode peak gain;

decode a shape of predetermined frequency bins surrounding the peak; and

scale the decoded shape by the decoded peak gain.

16. A user equipment (UE) comprising the decoder of claim **14**, said decoder configured to output the reconstructed transformed harmonic audio signal to further audio signal processing circuitry of the UE, for generating a corresponding audio signal.

* * * * *